REAL-TIME VIDEO ANALYSIS FOR INTRUSION DETECTION IN INDOOR ENVIRONMENTS

G. Milanesi *A. Sarti* *S. Tubaro*
Dip. di Elettronica e Informazione – Politecnico di Milano,
Piazza Leonardo Da Vinci 32, 20133 Milano, Italy
e-mail: Augusto.Sarti@polimi.it

ABSTRACT

In this paper we propose a novel system for indoor video surveillance. Our system, starting from a sequence of images, is able to detect and track moving objects even in the presence of significant variations of scene illumination. After a first analysis and clustering of the luminance time changes, a classification algorithm based on a fuzzy logic approach is used to identify moving regions that really represent unexpected objects moving in the scene, while discarding reflections and luminance profile changes due to illumination variations. One key feature of our system is its modest computation complexity, which allows it to operate in real-time on a standard PC platform. The real-time implementation of the system has been tested on a wide variety of situations, proving its effectiveness and robustness.

1 Introduction

In order to guarantee the necessary level of safety and effectiveness, an automatic video-surveillance system is expected to react to a wide range of complex situations correctly. This means that the system must be able to analyze what changed in the acquired images and recognize the presence of intruders. Typical scene chances that do not correspond to intrusions are changes in the environmental illumination, such as natural light dimming due to sudden clouding or sun setting; flickering of fluorescent tubes; lightbulbs switching on or off; car brights flashing through the windows; etc.). In addition, in order for the system to be able to corre ctly analyse regions of detected motion, shadows and reflexions due to moving objects should be detected and treated separately from the actual objects in motion.

In this paper we propose a novel intrusion detection system that is particularly suitable for indoor use. The system is able to detect and track moving objects that appear in the field of view of a static camera and is able to robustly distinguish between luminance profile changes due to a moving object and those due to illumination changes that can normally occur in the environment. The analysis system consists of three cascaded blocks:

Change detection and pre-classification: low-level analysis to extract the regions of change, and roughly distinguish between illumination changes and geometrical scene changes.

Attention focusing: temporal tracking of the regions of interest (bounding boxes) in order to regularise them and improve the preliminary classification performed by the previous block.

Classification: decision on whether the detected change in a certain area was, in fact, a geometric change (real intrusion), or a variation in the illumination, a shadow, a reflection, etc.

This three-block subdivision of the global approach makes our solution scalable because, even if we remove the last or the last two blocks from the chain, the reduced system will still be usable (with reduced performance) for intrusion detection purposes.

In the following three Sections, we will describe the three basic blocks of the system. Section 5 will provide more information on the global complexity of the system and present the results a series of tests conducted on real sequences.

2 Change detection and pre-classification

The first block of our video-surveillance system is aimed at an accurate frame-by-frame detection of the areas that exhibit significant changes between the current frame and previous frames (or some reference frame). In order to reduce the computational load, only changes in the luminance profile are considered. The algorithm, however, is designed in such a way to be relatively insensitive to changes in the global scene illumination.

The first step is to compute the difference between frames in order to localize those areas where significant luminance changes took place. There are several ways to do so, one very simple solution that takes into account both the differential change ($F_c - F_p$) between current frame and previous one, and the absolute change ($F_c - F_b$) between current frame and a reference one, consists

of computing pixel-by-pixel the map

$$\max\left[(F_c - F_p), (F_c - F_b)\right] \quad . \tag{1}$$

Here the reference frame (background) is a reasonably recent frame acquired knowing that there was no motion in the scene. If (1) exceeds a threshold T_h, then the corresponding pixel is labeled as "change point". At the end of this process we have a boolean mask M that specifies the presence of local changes. The threshold T_h is dynamically computed taking into account the local average and standard deviation of the samples produced by the map (1), in accordance with what proposed by Hamadami [3]. Pixels of the background frame F_b are not updated all in the same way. In fact, the update is faster for the pixels below threshold and slower for the others. This way it is still possible to keep track of small luminance changes that occur between frames, while structural scene changes (large objects that move) will not be treated like change areas for long.

The information contained in the change mask M is then improved through morphological closing, and the areas of interest (connected regions of change) are enclosed in bounding boxes. The result will be a set of partially overlapping rectangles which are finally fused together into a smaller number of larger non-overlapping boxes.

At this point we can limit our analysis to the detected bounding boxes in order to reduce the computational cost, and apply a robust algorithm that exhibits little sensitivity to luminance changes [1, 2]. If we adopt a simple multiplicative model for the scene illumination, then the luminance profile $F(x, y)$ will be the product between an "illumination profile" $I(x, y)$, which is assumed as slowly varying, and a "local texturing" $S(x, y)$, whose frequency content is more in the high range. As both reference frame $F_r(x, y)$ and current frame $F_c(x, y)$ are modeled as such a product

$$\begin{aligned} F_r(x, y) &= I_r(x, y)S_r(x, y) \\ F_c(x, y) &= I_c(x, y)S_c(x, y) \quad , \end{aligned} \tag{2}$$

the behavior of the ratio F_c/F_r will exhibit different frequency content, depending on what is changing in the scene. If the local change is purely in the illumination, then the ratio F_c/F_r will correspond to I_c/I_r, therefore its frequency content will be in the low range. Conversely, if the change is purely geometric, then the ratio F_c/F_r will correspond to S_c/S_r, therefore it will be rapidly varying. This can be easily exploited using two filters that extract the two frequency components of interest, like the local average $m(x, y)$ (low-pass) and the local standard deviation $\sigma(x, y)$ (high-pass). The analysis of this information is, at this point, quite straightforward:

- $|m| \simeq 1$ and modest σ imply that no change occurred between F_r and F_c;
- $|m| >> 1$ and modest σ imply that there has been a diffuse change in the scene, which is likely to be due to the illumination;
- large σ implies that there has been a signficant variation in the local texturing, which is likely to be due to geometrical changes in the scene.

Figure 1: Example of the variance map generated by the change detector. Darker regions correspond to a larger variance.

Indeed, there are many exceptions to this criterion, which are due to model failure. A multiplicative model of the illumination is, in fact, quite simplicistic, and is easy to fail, for example, in the presence of reflective surfaces and non-diffuse illumination. However, for matte surfaces and diffuse illumination, it performs quite nicely.

A quantized version of the local variance map M_v is stored by the system for future analysis (see Fig. 1).

3 Attention focusing and object tracking

The attention focusing block performs object tracking, as it searches for correspondences between bounding boxes in consecutive frames. This is done by seeking temporal continuity in both shape and motion. The tracking phase, as well as the classification phase, are based on the variance mask M_v generated by the previous block.

The tracking algorithm is based on the method proposed by Chetverikov [4]. It looks at three consecutive frames and determines the correspondence between bounding boxes through the minimization of an appropriate cost function f. This cost function takes into account both motion compatibility (based on the motion of the centroids of the variance mask within the considered boxes) and shape compatibility (based on the zero-order moment of the variance masks).

This solution is characterized by a good computational efficiency, particularly in situations like ours, where the number of change areas typically no more than 6 or 7.

4 Final classification

Once determined and tracked the areas where changes occurred, we need to classify them according to their origin. In particular, we want to distinguish geometric changes (moving intruders, sometimes only partially visible) from any other type of change (typically reflexions, shadows, and noise sources of other nature). The parameters used to discriminate between such two categories are:

- The ratio R between the zero-order moment of the variance mask and its perimeter. This parameter describes the "activity" of the luminance profile, after normalization on the part of the perimeter. This normalization action tends to make the parameter less sensitive to the distance from the viewpoint.

- The morphological index. This parameter is based on the so-called morphological spectrum [5, 6], which is an operator that extracts the contribution of every structural element from an image through a series of operations of morphological opening

$$f(n) = \frac{m(\Psi_n(M_v)) - m(\Psi_{n+1}(M_v))}{m(M_v)} \quad , \qquad (3)$$

  where Ψ_n is the morphological opening operator; m is the operator that computes the zero-order moment; and n is the size of the morphological opening's kernel (structuring element). The morphological spectrum of order n represents the contribution of the kernel n to the variance mask M_v. In what follows we will use a morphological index that incorporates the information contained in several morphological spectrum coefficients.

In order to characterize such parameters, we run a series of tests with various types of intruders (completely visible, partially occluded, etc.) and scenes (strongly changing lighting conditions, presence of reflections, etc.). We noticed that the variance masks M_v associated to intruders, are better described by morphological kernels of significant size ("relevant details") while non-geometric changes usually excite smaller kernels ("irrelevant details"). The morphological index, however, was determined through a joint statistical analysis of two morphological spectra of consecutive order. As we can see in Fig. 2, we notice that a good discrimination boundary is the dashed line at 45 degrees, which means that a good discriminant of the presence of intruders in a scene could be the difference of two morphological spectra of consecutive order. In Fig. 3 we can see a comparison between ratio R and the morphological index. The parameter R associated to an intruder turned out to be always significantly higher than in the other cases, therefore it represents a strong discriminant for the purpose of intrusion detection.

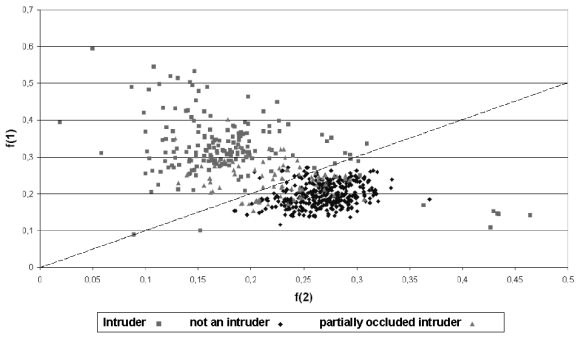

Figure 2: Distribution of the morphological spectra of orders 2 and 3, according to type of content (intruder, not an intruder, partially occluded intruder).

Now that we have a pair of good discriminants, we can use them jointly through a properly defined classifier. Our approach to this problem is based on fuzzy logic [7] and the semantic rules used for classification are:

- *IF R is High AND Morphological Index is Relevant THEN geometric (human) intrusion;*

- *IF R is Low AND Morphological Index is Irrelevant THEN non-geometric intrusion.*

The membership function relative to the input linguistic variables were determined through statistical analysis (hystograms) of the available data.

5 Performance Evaluation

The system performance was measured in terms of wrong classifications. We acquired our grayscale videos using both a low-quality webcam and a digital camera of good quality. The intruder, once in the scene, was allowed to change posture or partially hide behind furniture. The scene lighting was often (and suddenly) changed during the video acquisition. The results of these experiments are colleted in the following table.

classification / content	Intruder	Not an intruder	Uncertain
Total visibility	175	10	8
Partial visibility	484	65	2
Not an intruder	20	206	0

In spite of the worst-case selection of testing videos, the percentage of correct classification is around 89%. If we had used only the ratio R, the percentage of success would have dropped of more than 10%. The column labeled as "uncertain" denotes the situations in which the classifier was unable to make a decision. In all considered videos, however, the intruder's trajectory was always uniquely identified.

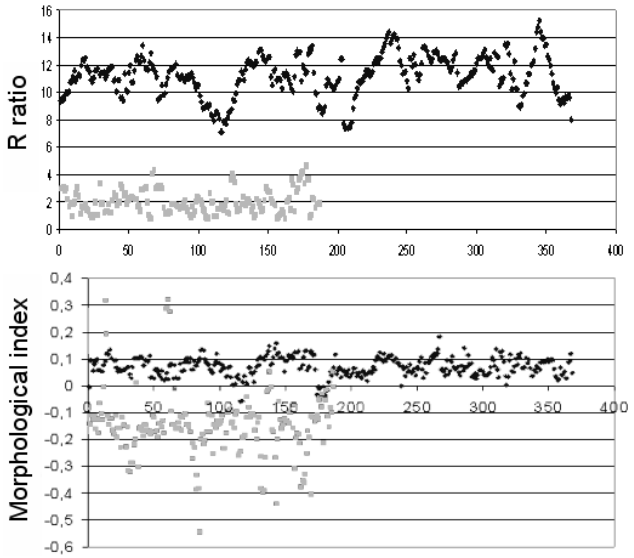

Figure 3: Intrusion discrimination based on the selected parameter. The light dots represent non-geometrical changes, while the black dots correspond to intruders. The abscissa corresponds to the scene index.

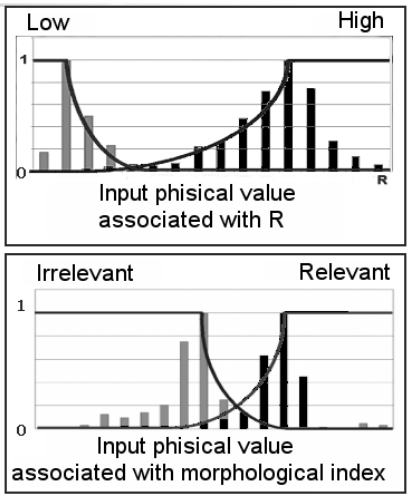

Figure 4: Membership function relative to the input linguistic variables of the fuzzy logic classifier.

Figure 5: Example of application. From left to right: original frame, output of the attention focusing block, output of the final classification block. In the top sequence the intruded is correctly detected already at the second stage of the algorithm. In bottom sequence a correct classification is only achieved with the third phase.

6 Conclusions

We proposed and implemented a video-surveillance system operating in real-time that turned out to be robust against illumination changes and shadows in the scene. The classifier proved able to correctly recognize intruders as such, even difficult acquisition conditions.

References

[1] T. Ebrahimi, E. Durucan, "Robust and Illumination Invariant Change Detection Based on Linear Dependence". *Proc. EUSIPCO 2000*, Tampere, Finland.

[2] K. Skifstad, R. Jain, "Illumination Independent Change Detection for Real World Image Sequences", *Proc. CVGIP*, Vol. 46, pp 387–399, 1989.

[3] G. X. Ritter, J. N. Wilson, *Handbook of Computer Vision Algorithms in Image Algebra*, CRC Press, 1996.

[4] D. Chetverikov, J. Verestoy, "Tracking Feature Points: a New Algorithm" *Proc. ICPR*, Vol. 2, pp. 1436–1439, 1998.

[5] P. Maragos "Pattern Spectrum and Multiscale Shape Representation", *IEEE Tr. PAMI*, Vol. 2, pp. 701–716, 1989.

[6] A. Asano "Texture Analysis Using Morphological Pattern Spectrum and Optimization of Structuring Elements" *Proc. ICIAP*, pp. 209–214, 1999.

[7] L. A. Zadeh "From Computing With Numbers to Computing with Words. From Manipulation of Measurements to Manipulation of Perceptions". *IEEE Tr. CAS I: Fund. Theory and Appl.*, Vol. 46, pp. 105–119, 1999.